DOI:10.31557/APJCC.2026.11.4.613

REVIEW

Artificial Intelligence and Precision Dermatology across Inflammation, Cancer, and Infection: A Review

Ihab Zaqout[1]*, Saeb Zakout[2]

[1]Highest Academic Rank/academic Degree (Professor), Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University of Gaza, Gaza Strip, Palestine; School of Electronics and Computer Science, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, UK. [2]Highest Academic Rank/academic Degree (Dermatologist), Zakout hudklinikk AS, Kristiansand, Norway.

Abstract

Introduction: Artificial intelligence (AI) is rapidly transforming dermatology, with emerging roles in the triage, diagnosis, and monitoring of inflammatory, neoplastic, and infectious skin diseases. Translation to routine care is limited by dataset imbalance, domain shift, and inconsistent external validation, calibration, and fairness reporting. The aim of this study was to review current applications of AI in dermatology, evaluate the performance of AI systems across selected inflammatory, neoplastic, and infectious skin diseases, and identify key challenges and priorities for their safe, equitable, and effective clinical implementation. **Materials and Methods:** This narrative review synthesizes recent evidence on AI workflows for eczema and psoriasis severity scoring (EASI/PASI), pigmented tumors (melanoma, basal cell carcinoma, naevi, and benign keratosis-like lesions), and selected infections (superficial mycoses and mpox). We summarize commonly used public, metadata-rich datasets (e.g., HAM10000, ISIC challenges, BCN20000, SIIM-ISIC 2020, PAD-UFES-20, Derm7pt) and methodological trends including convolutional and transformer backbones, detection and segmentation models, label-efficient pretraining, and multimodal/vision-language systems. **Results:** Across disease groups, the most mature evidence supports pigmented-lesion classification, where discrimination is high on curated dermoscopy benchmarks and clinician-AI collaboration improves triage decisions. Segmentation-derived EASI and PASI estimates show strong correlation with clinician assessments for longitudinal monitoring. For superficial fungal infections and mpox, image-based models are approaching dermatologist-level performance, but generalization remains sensitive to capture device, setting, and population. **Conclusion:** Key priorities for equitable and reliable precision dermatology include leakage-safe, skin-tone-balanced datasets, routine external and temporal validation with subgroup analyses, calibrated and uncertainty-aware outputs (including selective referral strategies), privacy-preserving deployment (e.g., on-device inference and federated learning), and adherence to standardized health-AI reporting frameworks.

Keywords: Artificial intelligence- deep learning- precision medicine- dermatology- skin disease

Asian Pac J Cancer Care, **11 (4)**, 613-629 Submission Date: 03/11/2026 Acceptance Date: 05/01/2026

Introduction

Artificial intelligence (AI) systems are currently achieving or surpassing expert-level performance in various medical fields; however, there are still unresolved questions regarding the generalizability of these models, their fairness across diverse populations, and the safety of their integration into real-world healthcare settings. In the field of dermatology, high-resolution imaging and multimodal data streams provide valuable visual and quantitative information for diagnosis, tracking disease severity, and making therapeutic decisions. Nevertheless, numerous datasets inadequately represent darker skin phototypes and other marginalized groups, which hinders equitable implementation [1]. Concurrently, image- and microscopy-assisted methodologies have demonstrated that specialized models can effectively differentiate between neoplastic, inflammatory, and infectious skin conditions when trained and validated on well-curated cohorts [2 - 4]. To transition from promising prototypes to safe and clinically applicable systems, regulatory bodies and expert organizations emphasize the importance of

Corresponding Author:
Dr. Ihab Zaqout
Highest academic rank/academic degree (Prfoessor), Department of Information Technology, Faculty of Engineering and Information Technology,Al-Azhar University of Gaza, Gaza Strip, Palestine.
Email: i.zaqout@alazhar.edu.ps; I.Zaqout@soton.ac.uk

transparent reporting, external validation, and careful claims to ensure that performance metrics can be translated into dependable point-of-care tools [5].

Precision dermatology seeks to provide the appropriate intervention to the right patient at the right moment by considering the variability in disease biology, comorbidities, environmental factors, and treatment history. In practice, this objective relies on the integration of routinely collected data types: clinical photographs, dermoscopy, histopathology, laboratory results, genomics, and longitudinal electronic health records, into cohesive patient trajectories. AI techniques facilitate this process by converting raw pixels and time-series data into structured phenotypes (such as diagnoses, endotypes, severity scores, and trajectories) and by developing prognostic and treatment-response models that can aid in individualized risk stratification and therapy selection, particularly for heterogeneous, relapsing inflammatory dermatoses.

A robust application of artificial intelligence in dermatology relies on publicly accessible or shareable, metadata-rich datasets that reflect variations in anatomy, disease progression, skin pigmentation, medical devices, and treatment environments. Multi-center image collections, such as PAD-UFES-20, regional skin-lesion cohorts (including datasets from Argentina and Europe), and established dermoscopy benchmarks like BCN20000, Derm7pt, and PH2 serve as foundations for supervised learning, highlight biases, and facilitate significant cross-site validation [6 - 13]. Strategies that preserve privacy, such as federated and split learning, de-identification, and synthetic data generation [14], can enhance the available training signals while ensuring the maintenance of consented provenance. Within this framework, rigorous "dataset hygiene" (including patient-level de-duplication, label auditing, and ensuring balance across skin tone, devices, and sites) is essential for reliable evaluation and comparison of AI systems in dermatology.

Generative (synthetic) data can also improve data efficiency by augmenting rare classes or under-represented skin tones, using methods such as GANs and diffusion models to create additional training images or to inpaint privacy-preserving variants. However, synthetic data can introduce subtle artifacts, amplify existing dataset biases, or leak memorized patient-specific patterns if models are trained or sampled improperly. In practice, synthetic images should be treated as augmentation rather than a substitute for real-world external validation, and their use should be documented with safeguards such as separation from test sets, clinician review for plausibility, and auditing for distribution shift.

Methodologically, the majority of contemporary dermatology models adhere to conventional image-classification workflows that refine convolutional or transformer architectures on dermoscopic and clinical images. These foundational architectures provide the necessary technical support for subsequent precision-dermatology applications [15 - 19]. Established segmentation frameworks, ranging from U-Net variants to promptable segmentation models, serve as prototypes for transforming pixel-level lesion masks into volumetric and extent measurements that can be correlated with clinical scores for conditions such as atopic dermatitis and psoriasis [20 - 27]. This technical groundwork is further enhanced by disease-specific pipelines that address particular clinical entities: AI for melanoma and non-melanoma skin cancers focusing on pigmented lesions and dermoscopic patterns [28, 29], deep learning systems targeting onychomycosis and other fungal nail infections [30 - 32], and image-based identification of mpox as a case study of emerging infectious dermatoses where swift visual triage is crucial [33].

Contemporary dermatology classifiers are progressively utilizing foundation models, including Vision Transformers and hybrid architectures, which are pre-trained on extensive generic or medical image datasets and subsequently fine-tuned for specific domain tasks [34 - 37]. Multimodal and vision-language frameworks further connect images with unstructured clinical notes, patient-reported outcomes, and decision-support outputs, facilitating joint reasoning across visual and contextual data [38 - 42]. In conjunction with these advancements, there is an increasing focus on calibration, coverage estimates, and fairness-aware evaluation, which includes audits for domain shifts, subgroup performance, and label noise. These factors highlight the importance of practical safeguards against data leakage, spurious correlations, and ambiguous cohort definitions when evaluating dermatology models [43 - 52].

Within various disease categories, the most developed evidence base is found in pigmented-lesion classifiers, which demonstrate high discrimination capabilities on curated dermoscopy datasets and are progressively integrating uncertainty quantification and referral strategies to enhance safe triage [53]. Simultaneously, segmentation-to-burden pipelines for eczema and psoriasis exemplify how pixel-level outputs can be synthesized into clinically meaningful severity indices, while infection-focused classifiers illustrate how microscopy and clinical images can aid in the identification of common pathogens and inform treatment decisions in routine clinical environments. Collectively, these examples demonstrate how dermatology AI can evolve beyond mere "spot the lesion" tasks towards comprehensive tools that connect morphology, burden, and risk over time.

Recent evaluations consolidate the applications of AI in inflammatory dermatoses, skin cancer, and infectious skin diseases, emphasizing the potential for precision medicine while also identifying ongoing deficiencies in real-world validation, health equity assessment, and regulatory alignment [54 - 60]. Expanding upon this body of work, the current review examines the interplay between technical decisions in data curation, model design, and evaluation with clinical contexts related to inflammation, cancer, and infection, and how these decisions can advance AI towards achieving equitable and reliable precision dermatology.

Skin Disease Categories

All of the following subsections outline clinical

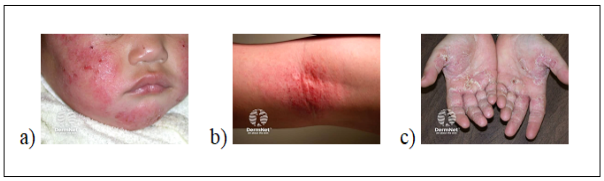

Figure 1. Eczema. (a) slightly raised red patches on the cheeks and around the eyes, showing mild facial eczema, (b) Red, thickened skin in the elbow creases with scratch marks, showing long-standing flexural eczema, (c) Red, dry, cracked skin on the hands, showing chronic hand eczema with fissures

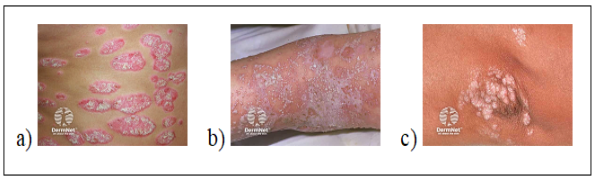

Figure 2. Psoriasis. (a) Well-defined red plaques with silvery scale on the limbs and trunk, showing typical chronic plaque psoriasis, (b) Large, clearly outlined pink plaques with silvery scale on the leg, showing extensive plaque psoriasis, (c) Acute psoriasis with small drop-like or pustular spots, often after an infection, showing guttate or pustular psoriasis

features for each category. Sources are provided for Eczema, Psoriasis, Melanoma, Basal Cell Carcinoma (BCC), Melanocytic Nevi (NV), Benign keratosis-like lesions (BKL), Seborrhoeic keratoses and other benign tumors, Tinea (ringworm), Candidiasis/other fungal infections, and Warts/Molluscum/other viral infections. Figs. 1-9 present representative images from a widely used public clinical dataset [61], randomly selected for illustrative and methodological purposes, and panels illustrate intra-class variability and common mimics that challenge image-only triage. Images are provided for illustrative purposes and are not intended for diagnostic use.

Eczema

Atopic dermatitis is a chronic, relapsing form of itchy eczema, and its severity is most accurately evaluated by integrating photographic images with clinical context, which encompasses factors such as age, location, symptoms, medical history, and associated atopic conditions. The distribution of the condition varies with age, and potential triggers include irritants, allergens, infections, and stress. Complications may arise, including secondary infections, eczema herpeticum, and erythroderma [31]. The EASI scoring system offers a standardized approach for assessing severity in both clinical practice and research trials [32]. Recent advancements in image-based AI techniques enable the estimation of atopic dermatitis severity scores, such as ASCORAD (an automated derivative of SCORAD), directly from clinical photographs through lesion segmentation and region-specific scoring [30]. Consequently, the automated evaluation of eczema severity represents a burgeoning area of AI application, featuring systems designed for lesion segmentation, severity assessment, and remote image-based monitoring, which facilitate more accurate longitudinal patient care [55]. Figure 1 depicts the range from mild facial rashes to chronic lichenified and fissured lesions.

Psoriasis

Psoriasis is a chronic immune-mediated condition characterized by distinct red scaly plaques and notable features such as the Auspitz sign, the Koebner response, and nail alterations that may signal a risk for psoriatic arthritis. The clinical manifestations range from plaque to guttate, inverse, pustular, and erythrodermic forms, with variations in presentation based on skin tone, where redness may appear violaceous and dyspigmentation can obscure certain characteristics. Segmentation-based Psoriasis Area and Severity Index (PASI) estimation, utilizing psoriasis-specific methodologies and advanced lesion segmentation architectures [21], aligns closely with clinician assessments and provides objective monitoring [27]. Figure 2 depicts the continuum from stable chronic plaques to abrupt widespread outbreaks. In addition to traditional segmentation-based PASI methodologies, end-to-end convolutional neural network frameworks are now capable of estimating PASI directly from multi-view clinical images, demonstrating a high level of agreement with clinician evaluations [62]. Furthermore, specialized reviews compile workflows for image-based psoriasis analysis and highlight ongoing deployment challenges [54].

Melanoma

Melanoma represents a potentially fatal tumor of melanocytes, necessitating early detection and surgical intervention. Clinical guidelines such as ABCDE and EFG, along with an understanding of acral and subungual sub-types, are crucial for assessment, particularly in individuals with darker skin tones [guideline/ref]. Foundational AI systems in dermatology can achieve performance levels comparable to dermatologists in the classification of dermoscopic images, proving most effective when utilized in conjunction with clinical judgment [15, 17, 18, 63]. Current research is directed towards enhancing generalization and minimizing bias in ResNet-based classifiers that are trained on extensive public datasets, incorporating explicit subgroup and fairness analyses [64 - 66]. Additional studies combine clinical metadata with images to enhance discrimination and calibration [67]. Beyond image-based screening, deep learning models that are trained on whole-slide histopathology images have the capability to predict one-year disease-free survival in cutaneous melanoma, thereby aiding in risk stratification and treatment planning [68]. Figure 3 underscores the importance of identifying amelanotic and non-pigmented presentations that may resemble inflammatory skin diseases.

Basal Cell Carcinoma (BCC)

Basal cell carcinoma, commonly referred to as BCC, represents a frequently occurring skin cancer characterized by local destruction. The various subtypes and risk factors

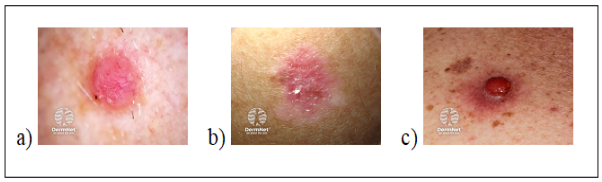

Figure 3. Melanoma. (a) Amelanotic nodular melanoma - a red, inflamed-looking lump, (b) Amelanotic scaly melanoma - a pink, scaly patch that can look like dermatitis, (c) Superficial amelanotic melanoma - a flat red patch on the skin

associated with BCC make it imperative to achieve accurate diagnosis in order to reduce morbidity [60]. While BCC is generally included in multiclass clinical and dermoscopic AI classifiers, these tools are meant to assist, and it remains crucial to obtain clinicopathologic confirmation in instances where presentations are unusual or inflammation obscures the margins [15]. Figure 4 demonstrates the spectrum of BCC appearances and emphasizes the diagnostic challenges encountered by healthcare professionals and automated image analysis tools.

Melanocytic Nevi (NV)

The differentiation of benign melanocytic nevi from lesions that necessitate dermoscopy monitoring or biopsy is based on a combination of clinical assessment, dermoscopy, and algorithmic tools. Segmentation frameworks [69] and classification models that focus on attributes for melanoma [15, 17, 18] enhance the accuracy of classification and the detection of melanoma. Incorporating clinical metadata such as age, site, and patient context leads to improved performance, while fairness studies stress the necessity for diverse representation of skin tones [65 - 67]. Common acquired nevi are typically symmetrical, with uniform color and smooth edges, while congenital nevi are larger and exhibit greater color and texture variation; thus, transient irritation or infection should trigger a reevaluation after healing. Figure 5 illustrates how temporary inflammatory changes can resemble pathology, underscoring the need for careful clinical and algorithmic distinction.

Benign keratosis-like Lesions (BKL)

Identifying dermoscopic indicators such as milia-like cysts, comedo-like openings, and moth-eaten borders facilitates the differentiation of benign keratosis-like lesions from melanoma, thereby aiding in annotation and clinical decision-making [70]. Given that crusting, inflammation, ulceration, bleeding, or rapid changes can obscure benign characteristics, lesions exhibiting these alterations or those located in atypical areas, such as the ear, should be subjected to clinicopathologic verification. Figure 6 underscores the necessity of incorporating inflamed and atypical cases in training and evaluation datasets to minimize misclassification by automated systems.

Seborrhoeic Keratoses and Other Benign Tumours

Identifying the clinical and dermoscopic characteristics of seborrhoeic keratoses and prevalent benign tumors minimizes false positives and enhances triage in AI-driven workflows. Multiclass AI can attain a level of discrimination comparable to that of dermatologists, and training on extensive, diverse datasets, as well as community and clinical corpora, fosters improved generalization. Incorporating clinical metadata, such as age and location, into ResNet pipelines improves calibration and reduces misclassification [65 - 67]. Nevertheless, in pigmented lesion benchmarks, seborrhoeic keratoses (SK) and benign keratosis lesions (BKL) frequently exhibit lower accuracy per class compared to nevi (NV) and basal cell carcinoma (BCC), underscoring the necessity for well-labeled examples [15]. This underperformance likely reflects a combination of factors: BKL is a heterogeneous catch-all label that includes visually diverse entities, its dermoscopic patterns can overlap with early melanomas and other malignancies, and training labels may be noisier when lesions fall near borderline diagnostic categories or are confirmed by variable reference standards. Improving BKL performance therefore often requires more granular labeling, higher-quality ground truth, and targeted sampling of hard negatives.

Distinct morphological indicators, such as milia-like cysts and comedo-like openings in SK [70], along with

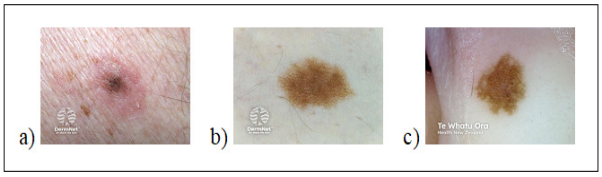

Figure 5. Melanocytic Nevi. (a) Melanocytic nevus with mild eczematous halo, looks inflamed but benign, (b) Slight redness or irritation, mimicking inflamed skin, (c) Nevus on foot/trunk, slight irritation from friction, subtle inflammatory look

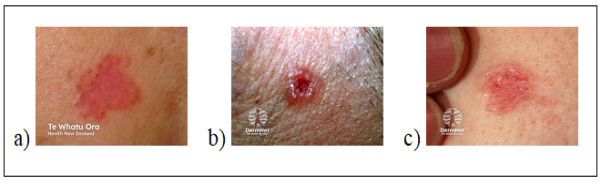

Figure 4. Basal Cell Carcinoma. (a) Pink plaque with indistinct margin, mimicking chronic dermatitis, (b) Pink, scaly/crusted plaque on face, mimics eczema/rosacea, (c) Small pink patch with minimal classic BCC features, mimics purely inflammatory lesion

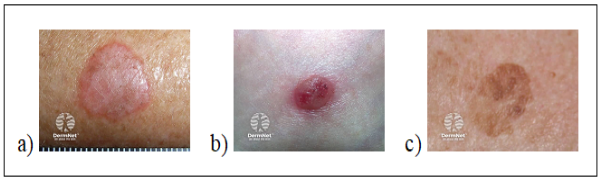

Figure 6. Benign keratosis-like lesion. (a) Red/pink macule or thin plaque inflamed but benign, (b) Seborrhoeic keratosis with red halo, scale or crust, can mimic dermatitis, (c) Classic keratosis plaque with subtle surrounding erythema/irritation

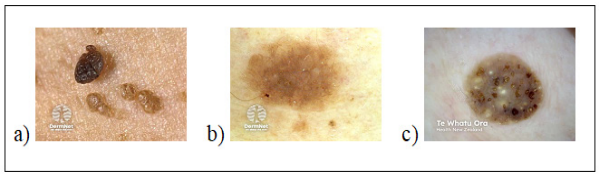

Figure 7. Seborrhoeic keratoses & other benign tumors. (a) Typical "stuck-on" brown waxy plaque, (b) Benign keratosis showing erythema/irritation - benign but looks inflamed, (c) A benign keratosis with more complex surface morphology

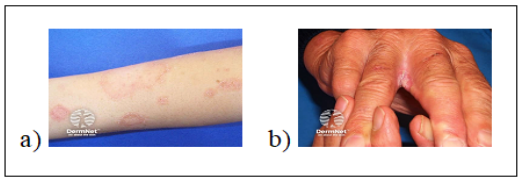

Figure 8. Tinea / Ringworm / Candidiasis. (a) Tinea corporis (ringworm on the body), (b) Candida intertrigo (yeast infection in skin fold)

classic signs like the dimple sign in dermatofibromas, crown vessels in sebaceous hyperplasia, and lacunar patterns in angiomas, offer dependable training labels. Given that inflammation can obscure these patterns, atypical or persistent lesions necessitate dermoscopy-guided biopsy. Figure 7 emphasizes the significance of incorporating both classic and irritated keratotic appearances in training and validation datasets to mitigate false positives when identifying inflammatory or malignant lesions.

Tinea, Candidiasis and Other Fungal Infections

AI-assisted fluorescence microscopy analyzers for superficial fungal infections have been suggested as preliminary diagnostic instruments and may be incorporated into image-based triage processes [4]. Automated and AI-enhanced techniques, including prospective photo-based models for onychomycosis and broader multi-disease classifiers, can achieve accuracy comparable to that of dermatologists and facilitate point-of-care triage [2, 3]. Clinicians must remain vigilant regarding steroid-modified or travel-related presentations, as well as emerging species such as Trichophyton indotineae, since these factors can influence management strategies and model training. Figure 8 and site-specific image panels offer essential training data to enhance real-world classification and clinical triage.

Warts, Molluscum and other Viral Infections

Identifying the distinct clinical and dermoscopic characteristics of cutaneous viral infections enhances bedside diagnosis and aids in differentiating these conditions from inflammatory or neoplastic lesions. This differentiation, in turn, facilitates precise photo-triage and model training. Curated image collections serve as essential training data and, when integrated with multi-disease AI classifiers for viral exanthems, especially mpox, could allow for efficient front-door triage of viral skin diseases [71, 72]. Well-annotated examples of prevalent cutaneous viral eruptions are illustrated in Figure 9.

Cross-cutting AI Methods

This section unifies the core AI pipeline as depicted in Figure 10. It begins with defining the clinical question (triage/diagnosis vs. severity monitoring), followed by data acquisition and curation to ensure balanced skin tones/devices, leakage-safe splits, and external validation. Learning then proceeds with modern vision backbones (CNN/ViT, pretrained) and optional multimodal fusion of images with metadata. Prediction spans image-level classification, detection, and pixel-level segmentation, producing clinically meaningful outputs such as EASI/PASI components, lesion extent, trajectories, and flare detection.

In clinical workflows, these tools are best viewed as upstream decision support within a sequential pathway: patient or primary-care image capture → AI risk/triage output → clinician review (history, dermoscopy where available) → biopsy for concerning lesions → histopathology confirmation and management. When calibrated to prioritize sensitivity, AI triage can increase the pre-test probability of malignancy among referred/biopsied cases (enriching the queue), which may reduce unnecessary urgent referrals and focus dermatopathology time on higher-risk lesions. Conversely, overly low thresholds may increase biopsy volume and downstream pathologist workload, highlighting the importance of

Table 1. Comparison of Common Vision Backbones Used in Dermatology AI

Backbone	Typical performance	Data needs	Computing cost	Interpret-ability	Common uses
CNN (ResNet/ EfficientNet/ ConvNeXt)	Strong baseline; efficient for textures/ local patterns; often stable with moderate augmentation.	Moderate; good with thousands to tens of thousands of labeled images; benefits from pretraining.	Low-moderate; fast training/inference; ideal for mobile/edge compression.	Moderate; saliency helps inspection but is not inherently explanatory; needs calibration/ uncertainty.	Pigmented-lesion classification; inflammatory triage; fungal/viral image classification; on-device screening.
ViT (ViT/Swin)	Very strong with foundation pretraining; global context can help subtle or wide-field patterns.	High; typically, more pretraining/ data dependent on peak performance.	Moderate-high; attention increases compute/memory; optimized variants mitigate cost.	Low-moderate; attention maps can support review but do not guarantee interpretability.	Foundation-model fine-tuning; detection/ segmentation backbones; multimodal alignment.
Hybrid (CNN-Transformer; UNet+Transformer)	Balances local detail + global context; often strong for segmentation and structured severity outputs.	Moderate-high; can be efficient when initialized from pretrained components.	Moderate; design dependent; often heavier than pure CNNs.	Moderate; masks can be clinically interpretable, while feature-level reasoning remains complex.	Lesion segmentation; PASI/EASI pipelines; cascades; WSI patch screening + aggregation.

Table 2. AI Tasks, Methods, and Diagnostic Performance Across Precision Dermatology Applications

Disease / Theme	Task / Role	Input & Context	Performance Metrics	Reference
Atopic dermatitis (eczema) and psoriasis	Segmentation → EASI / PASI severity scoring	Clinical photographs of eczema and psoriatic plaques (often trunk images)	≈25% relative improvement in lesion-segmentation F1/precision; erythema accuracy 99.17%; PASI area scoring accuracy up to 79.4%.	[27, 30]
Melanoma / pigmented lesions - benchmarks, multiclass, domain shift	Image-level classification (single-lesion dermoscopy & clinical photos)	Dermoscopy datasets (e.g. HAM10000) and clinical photographs (e.g. BCN20000); dermoscopic multiclass pigmented-lesion benchmarks	AUC ≈0.98 on dermoscopy. Under dermoscopy→clinical-photo shift, AUC 0.981→0.907 and accuracy 82.0%→58.8%. Multiclass accuracies: BCC ≈0.91, NV ≈0.76, BKL ≈0.43.	[15, 90, 99]
Melanoma - decision support and primary-care triage	Reader-AI decision support and prospective triage	Dermoscopy images with dermatologist comparison; clinical photos from primary care	Reader-AI: sensitivity 87.5% at 60% specificity; at dermatologist sensitivity 74.1%, CNN specificity 86.5%. Primary care: AUC 0.960 (95% CI 0.928-0.980), sensitivity 95.2%, specificity 84.5%.	[18, 29]
Mpox and infectious exanthems	Image-level classification; lesion detection; smartphone deployment	Web-scraped and clinical photos of mpox-like lesions; public mpox image dataset; smartphone photos via mobile app	Pooled mpox AUC ≈0.97; sensitivity 83-91%, specificity 90-97%, prospective sensitivity ≈89%. Public dataset: 99.49% accuracy. Mobile app: 91.11% accuracy.	[33, 71, 72, 82, 91–94]
Superficial fungal infections and onychomycosis	Microscopy and photo-based nail classification	Fluorescence microscopic images; clinical / dermoscopic nail images	Superficial fungal infections: sensitivity 96.27%, specificity 94.92%, accuracy 96.00%. Onychomycosis: AUC 0.90-0.997; sensitivity 72-96%, specificity 76-96%; prospective AUC 0.751 with 70.2% sensitivity, 72.7% specificity.	[2 - 4]
Generic lesion detection (all diseases)	canonical COCO performance (non-dermatology)	COCO-like natural images, conceptually adapted to wide-field clinical photographs	RetinaNet-101-FPN (COCO, 800 px): AP 39.1. YOLOv4: AP 43.5, AP50 65.7% at ~65 FPS.	[75, 76]
Multimodal and vision-language models	Image-level classification with clinical context; classification, retrieval, and VQA	Dermoscopy / clinical photos plus metadata (age, sex, site, symptoms); images + free-text prompts or descriptions	Image + metadata and VLMs report accuracy, AUC, sensitivity, specificity, precision, recall, F1; retrieval uses Recall@K, mAP; VQA uses EM, BLEU, ROUGE plus clinician-rated correctness/safety.	[7 - 9, 12, 40]
Self-supervised and weakly supervised pretraining	Label-efficient representation learning	Large pools of unlabeled or weakly labeled generic + dermatology images	Pretraining (MAE, DINOv2, SimCLR, CLIP, BiomedCLIP) improves accuracy and AUC with few labels; downstream tasks report accuracy, AUC, sensitivity, specificity, precision, recall, F1.	[38 - 41]
Dataset design and reporting frameworks	Data curation, splitting, and guideline-aligned reporting	Dermatology image corpora with metadata (skin tone, device, site) across multiple diseases	Leakage-safe, tone/device-balanced splits aim to keep sensitivity, specificity, and AUC stable on external tests; guidelines recommend reporting accuracy, sensitivity, specificity, AUC, calibration (ECE, Brier), F1, and subgroup results.	[7 - 9, 43 - 46]
Explainability, uncertainty, and human-AI teaming	Saliency, calibration, selective prediction, and assisted diagnosis	Same inputs as diagnostic classifiers; psoriasis, eczema, mpox, and skin cancer datasets with saliency and human-reader studies	Saliency + deep ensembles / MC dropout improves calibration (ECE, Brier) and error detection while tracking accuracy and AUC; conformal prediction gives 90-95% coverage with reported sensitivity, specificity, F1; human-AI teams outperform clinicians alone.	[5, 74, 83, 85, 86, 97]

threshold selection and prospective workflow evaluation.

Table 1 provides a concise, practical comparison of the most common computer-vision backbone families used in dermatology AI (CNNs), Vision Transformers (ViTs), and hybrid models. It summarizes how these architectures typically differ in performance, data requirements, computational cost, and interpretability, and links these trade-offs to real dermatology use cases (e.g., lesion classification, segmentation, and point-of-care triage). This overview is intended to help readers select an appropriate model family based on task constraints such as available training data, need for real-time inference, and the level of explainability required for clinical deployment.

By Clinical Question

Image-level diagnosis and triage generally refine ConvNeXt, ViT/Swin, or EfficientNet backbones by incorporating class-imbalance management (such as focal loss) and post-hoc calibration, allowing for the customization of thresholds for screening compared

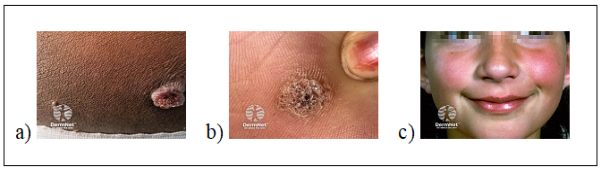

Figure 9. Viral Eruption. (a) Many small, smooth, skin-coloured bumps with a tiny central dip, (b) Thick, rough papules or plaques, sometimes with small black dots, (c) Bright red rash on the cheeks ("slapped cheek"), sometimes spreading to the arms and torso

to confirmatory applications. A prevalent example is a psoriasis-eczema triage model developed using clinical images, where a calibrated ConvNeXt classifier is fine-tuned to achieve high sensitivity to inflammatory lesions in primary care settings, while a more specific operating point is designated for teledermatology evaluations or in-person specialist consultations [34 - 37, 73, 74]. This context highlights that the clinical inquiry whether it be screening, triage, or differential support, dictates the selection of operating points, even when the foundational architecture remains consistent. Typically, pipelines incorporate robust data augmentation, optional ensembling, and test-time augmentation to enhance prediction stability without necessitating extensive metadata.

Label-efficient and resilient triage models are increasingly dependent on self-supervised or multimodal pretraining techniques. In the case of psoriasis, for example, a classifier that is initially pre-trained using MAE on millions of generic images and subsequently fine-tuned on plaque photographs can leverage more sophisticated visual features, resulting in improved performance in difficult areas such as the scalp or intertriginous regions. This demonstrates how MAE, DINOv2, SimCLR, CLIP, and multimodal BiomedCLIP contribute to enhanced representation quality when dermatological labels are scarce [38 - 42].

For inquiries related to disease burden, pipelines that segment severity are essential. U-Net and SAM/MedSAM/SAM-2 are utilized to segment plaques or eczematous patches, followed by straightforward morphological post-processing that transforms these masks into EASI or PASI component scores. In the context of eczema, this facilitates the automated assessment of erythema and extent in each body region, thereby aiding in the longitudinal monitoring of flare-ups and treatment responses. In psoriasis, similar pipelines are employed to estimate PASI by evaluating plaque area along with erythema, induration, and scaling [20, 22 - 27].

In the case of wide-field images or total-body photography, one-stage detectors like RetinaNet and YOLOv4 serve as initial lesion detectors, suggesting candidate boxes for pigmented lesions, inflamed plaques, or vesiculopustular eruptions, including mpox-like lesions sourced from curated web and clinical datasets, before passing these candidates to subsequent classifiers [33, 71, 72, 75, 76]. Consequently, the choice of whether the primary architectural component is a classifier, detector, or segmenter is predominantly influenced by the nature of the clinical inquiry, whether it is "what is this?", "where are all the lesions?", or "how severe is the disease?"

By Data Modality

Core inputs consist of two-dimensional images, including clinical photographs and dermoscopy, which are trained using leakage-safe patient-level splits. These inputs are evaluated against external datasets such as HAM10000, PH2, ISIC, and BCN20000, specifically for pigmented lesions. In a standard scenario involving melanoma versus nevus, a dermoscopic classifier is developed using HAM10000 and subsequently validated externally on BCN20000 or a dataset from a local hospital to assess performance amidst device and population shifts. The same foundational principles are applicable when utilizing backbone architectures for the detection of psoriasis plaques in smartphone images or when investigating image-based screening of mpox rashes sourced from social media or telehealth platforms, even if the specific datasets diverge from traditional skin-cancer collections [7, 11, 13, 15]. The integration of vision-language and multimodal methodologies further enhances this framework.

AI in Dermatopathology

AI is also increasingly used in dermatopathology, where the input is a digitized whole-slide image (WSI) rather than a clinical photograph. In the pathologist's workflow, WSI models can act as automated screeners that highlight suspicious regions, assist mitotic figure detection/mitotic counts, and quantify tumor-infiltrating lymphocytes and other microenvironment features. These quantitative outputs can support staging and risk stratification (e.g., in cutaneous melanoma) and can reduce time-to-review by prioritizing the most informative areas of a slide. Crucially, these systems are best positioned as decision support and workflow-acceleration tools; final interpretation and reporting remain with the dermatopathologist.

CLIP and the multimodal BiomedCLIP facilitate the mapping of images and text into a unified space, allowing for associations such as "erythematous scaly plaques on extensor surfaces" with psoriasis-like embeddings, and "vesiculopustular rash with umbilicated lesions" to underscore mpox-like characteristics without necessitating task-specific retraining [40, 41]. Assistant-style vision-language models (VLMs) like LLaVA-Med, Med-Flamingo, and SkinGPT-4 add conversational functionalities, enabling inquiries such as "show similar cases of flexural eczema in darker skin types" or "compare this plaque to classic chronic plaque psoriasis," providing both image retrieval and narrative explanations in response [77 - 79]. In summary, the modality dimension encompasses not only images but also clinical notes, symptom descriptions, and even patient-reported outcomes, all of which can be integrated to enhance precision dermatology tasks related to inflammatory diseases, cancer, and infections.

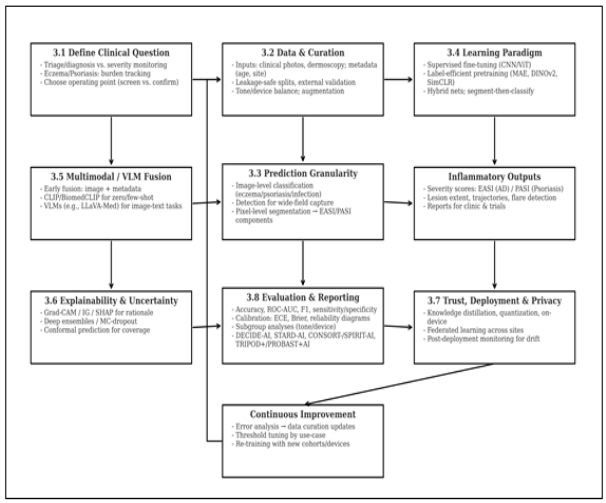

Figure 10. End-to-end AI Pipeline for Inflammatory Skin Disease

By Prediction Granularity

Image-level classifiers provide a singular label for each lesion or photograph, such as "psoriasis", "eczema", or "non-specific dermatitis". They are particularly effective for triage or high-level differential diagnosis when exact localization is not of utmost importance. In contrast, region-level detectors generate bounding boxes around potential lesions, allowing for the differentiation of a combination of mpox lesions and surrounding skin in wide-field images. This capability facilitates lesion counting and filtering prior to further analysis [75, 76]. Pixel-level segmenters create dense masks that can be transformed into measures of burden and severity. In the cases of psoriasis and eczema, segmentation models outline plaques or eczematous areas on trunk, flexural, and facial images, subsequently aggregating these into PASI- or EASI-like scores for ongoing monitoring [20, 22 - 27]. Therefore, the level of prediction granularity, ranging from image-level labels to region proposals and full-resolution masks, should align with the clinical objective, whether it is broad categorization, lesion counting, or precise severity assessment.

By Learning Paradigm

Most research predominantly depends on the supervised fine-tuning of CNN or ViT architecture utilizing Batch Normalization and Adam/AdamW optimizers, which are generally initialized from weights pretrained on ImageNet [64, 80, 81]. A notable instance is a classifier designed to differentiate between psoriasis, eczema, and other inflammatory dermatoses, constructed on a ViT backbone and fine-tuned using labeled outpatient photographs with augmentation techniques (including random crops, flips, and color jitter) to effectively represent real-world variations in lighting and pose [19]. Comparable supervised configurations are employed for melanoma and can be seamlessly adapted to other inflammatory and infectious diseases. Self-supervised pretraining approaches, such as MAE, DINOv2, and SimCLR, mitigate the reliance on extensive, fully labeled dermatology datasets by initially acquiring generic visual representations from unlabeled images, followed by task-specific fine-tuning [38, 39, 42].

For example, a substantial collection of unlabeled photographs of eczema and psoriasis can facilitate DINOv2 pretraining, after which relatively small, well-annotated datasets for eczema severity are adequate for training precise EASI scoring models. Vision-language models, including CLIP and the multimodal BiomedCLIP, further advance this paradigm by integrating text and images into a unified space. This integration facilitates zero-shot recognition of descriptors such as "umbilicated pustules on palm," which are associated with mpox-like exanthems, even in scenarios where task-specific data is scarce [40, 41].

Two recurring hybrid patterns are observed: (i) CNN-Transformer hybrids utilized for segmentation, where Transformer blocks offer global context for intricate patterns like overlapping psoriatic plaques or confluent eczema lesions, and (ii) segment-then-classify cascades, wherein an initial network isolates lesions (for instance, mpox pustules or eczema patches) and a subsequent network classifies or scores them, thereby minimizing background confounders and enhancing robustness [21, 69]. In practical applications, the learning strategy is tailored to the type of disease and the data regime: self-supervision is employed for limited mpox images, straightforward supervised fine-tuning is applied for prevalent psoriasis plaques, and hybrid segmentation-classification pipelines are utilized when accurate burden or severity estimation is necessary.

By Multimodal Fusion Strategy

Early/feature-level fusion integrates image embeddings with structured metadata, including age, sex, lesion site, and comorbidities. On datasets such as HAM10000, SIIM-ISIC 2020, PAD-UFES-20, and Derm7pt [7 - 9, 12], this approach consistently enhances discrimination and calibration compared to image-only baselines [57 - 59]. For instance, in the case of psoriasis, the combination of plaque images with patient age, disease duration, and known psoriatic arthritis status can aid in differentiating psoriasis from eczema or other papulosquamous eruptions when morphology alone is insufficient. In the context of mpox, incorporating travel history, contact exposure, or

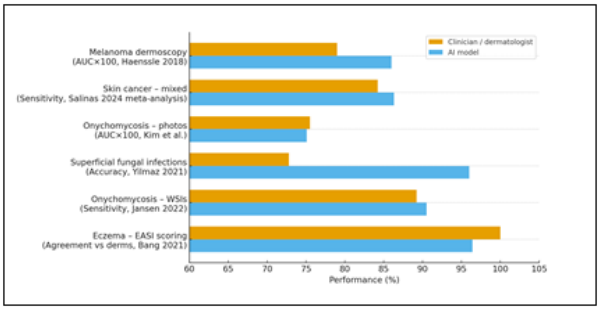

Figure 11. Sample Comparison of Clinician/ dermatologist and AI Performance Across Selected Skin Diseases, Showing that AI often Matches or Exceeds Clinician Performance

systemic symptoms alongside lesion photographs can refine triage decisions when distinguishing mpox from varicella or hand-foot-and-mouth disease, despite the fact that the underlying fusion patterns are analogous to those used in cancer and inflammatory tasks [33, 71, 72, 82]. Vision-language alignment techniques, such as CLIP and multimodal BiomedCLIP, facilitate the creation of joint representations of images and free text, enabling zero-shot labeling and concept-based retrieval [40, 41].

Assistant-style VLMs, including LLaVA-Med, Med-Flamingo, and SkinGPT-4, expand multimodal workflows beyond mere fusion: a clinician can present a photograph of flexural erythema with lichenification, pose the question, "does this resemble chronic eczema or inverse psoriasis?", and obtain both probabilities and a natural-language explanation that is grounded in the image [77 - 79]. Similar inquiries can be made regarding mpox ("does this rash pattern align with classic mpox?") or for intricate pigmented lesions accompanied by surrounding eczema changes. It is advisable to report subgroups by skin tone and device when utilizing such multimodal inputs, as metadata and language prompts may interact with image features in ways that differentially influence performance across Fitzpatrick types and camera sources [43].

By Causal/Decision Methodology

Methods of pair explainability, such as Grad-CAM [83], Integrated Gradients [84], and SHAP [85], are integrated with uncertainty-aware and calibrated predictions to facilitate safe clinical decision-making. In the context of psoriasis, Grad-CAM overlays can illustrate that a classifier concentrates on the erythematous scaly plaques instead of the surrounding skin or clothing, providing reassurance to clinicians that the model is utilizing disease-relevant indicators. In cases of eczema, saliency maps that emphasize flexural folds, lichenification, or excoriations can assist in distinguishing genuine disease involvement from incidental redness, thereby guiding more precise treatment decisions. For suspected mpox, explainability can reveal whether a model is focusing on characteristic umbilicated lesions as opposed to nonspecific erythema, which is particularly crucial during early outbreaks when clinical experience may be scarce. Uncertainty estimation through deep ensembles, Monte-Carlo dropout, or evidential networks enables models to identify low-confidence cases.

An underexplored but clinically important direction is correlating AI "pixel-level" explanations from clinical photographs with histopathology when biopsies are available. In many studies, saliency maps are assessed qualitatively by clinicians, but few link highlighted regions to co-registered microscopic features (e.g., whether attention within a psoriasis plaque corresponds to histological hallmarks such as parakeratosis with Munro microabscesses, elongated rete ridges, or dermal inflammatory infiltrate). Building such clinico-path correlation requires paired image–biopsy datasets, precise lesion localization, and workflows that allow dermatologists and dermatopathologists to jointly adjudicate whether model explanations align with biology rather than confounders (lighting, hair, scale, or background).

For instance, when plaques are atypical and may signify psoriasis, eczema, or drug eruption, uncertainty-aware models can flag these instances for referral rather than forcing a single confident label [86]. Calibration metrics such as ECE and Brier score quantify how well predicted probabilities match observed frequencies, which matter when thresholds trigger biopsy, antiviral treatment, or watchful waiting [74]. Conformal prediction can wrap a classifier to return a small set of plausible labels (e.g., "psoriasis OR eczema") with distribution-free coverage guarantees, making uncertainty explicit in triage and decision-support workflows [5].

By Trust & Deployment Constraints

Trust is founded on the integrity of datasets, which includes patient-level de-duplication, secure splits to prevent leakage, external validation, and subgroup analyses that consider variations in skin tone, device usage, and clinical environments. In the context of eczema and psoriasis, it is essential to ensure that images originating from the same patient or flare episode do not inadvertently cross over between training and testing datasets. Additionally, models should be assessed in new clinical settings or over different time frames to avoid exaggerating their effectiveness in managing chronic diseases. For mpox and other infectious diseases, it is vital to conduct external validation across various outbreaks or countries, because viral variants, underlying skin conditions, and photographic methodologies can vary significantly [33, 71, 72, 82]. Subgroup analyses hold particular significance for inflammatory and infectious diseases that manifest differently on darker skin, where signs of erythema may be less pronounced and terms like 'redness' may not be as visually apparent [65 - 67].

For high-stakes decisions, especially suspected melanoma, AI outputs should be treated as a highly sensitive screening or triage signal rather than a definitive diagnosis. In practical deployment, the operating point should prioritize sensitivity to minimize missed cancers, while the confirmatory step remains biopsy followed by histopathology. Point-of-care integration should also specify how and when inference occurs. In primary care or teledermatology, models may run in real time during image capture (with prompts for focus/lighting) or as an asynchronous post-hoc review that prioritizes cases in a queue. Outputs should be designed for actionability: a calibrated probability or risk category, an optional heat map/saliency overlay, and a short differential or 'refer/biopsy/monitor' recommendation linked to local pathways. Presenting uncertainty (e.g., low-confidence flags or abstentions) and logging the displayed result in the clinical record can improve transparency, auditability, and user trust. Medico-legal and liability considerations are an additional barrier to adoption. If an AI system fails to flag a melanoma, responsibility may be shared across the clinician, the deploying organization, and the manufacturer depending on jurisdiction, local policy, and whether the tool is used within its intended use. Practical

safeguards include clear governance (AI as decision support), user training, audit trails/documentation of AI outputs, monitoring for performance drift, and escalation rules that ensure clinical concern triggers referral/biopsy regardless of the model score.

The constraints of deployment also play a critical role in shaping model architecture. Knowledge distillation enables a robust, high-performing teacher model (for instance, a psoriasis severity estimator) to guide a smaller student network that is optimized for on-device inference, facilitating ongoing monitoring in community or home environments. Both knowledge distillation and federated learning facilitate on-device and multi-site training for tasks such as melanoma detection, psoriatic plaque identification, or mpox classification without the need to centralize raw images. Each participating center conducts local training and only shares updates on model weights, thereby safeguarding privacy while leveraging a more extensive collective dataset [87 - 89]. Collectively, these infrastructural decisions affect the speed at which models for eczema, psoriasis, and mpox can transition from research datasets to practical applications in clinical settings and teledermatology platforms, all while upholding trust and adhering to data protection regulations.

By Evaluation Design

Research employs leakage-safe, patient-level divisions with external or temporal validation and stratified subgroup reporting to align performance with the requirements of clinical deployment. For example, in the case of psoriasis, an EASI- or PASI-estimation model may be developed using images from a single tertiary center and subsequently evaluated on photographs from community dermatology or telehealth settings to assess the generalizability of severity scoring beyond idealized imaging conditions [20, 22 - 27]. Similarly, for eczema, evaluation frameworks should encompass subgroup analyses that consider skin tone, age, and care setting, thereby reflecting the diversity of real-world scenarios and ensuring that performance metrics are not exaggerated for underrepresented populations [65 - 67]. In the context of mpox, evaluations may specifically stratify by outbreak wave, geographic area, and co-infections to guarantee that models maintain their reliability as the viral epidemiology changes [33, 71, 72, 82].

It is also important to recognize that histopathology, while the clinical reference standard, is not a perfect "ground truth." Inter-observer variability among dermatopathologists is well documented, particularly for borderline melanocytic lesions (e.g., dysplastic nevi versus early melanoma) and for some inflammatory dermatoses with overlapping patterns. This label noise places an upper bound on achievable model performance and can lead to unstable training signals when datasets are small or heterogeneously annotated. Practical mitigations include multi-reader labeling with adjudication, consensus or probabilistic labels, reporting inter-rater agreement, and separating 'borderline/indeterminate' categories where clinical workflows already depend on multidisciplinary review.

Calibration metrics such as ECE, Brier score, and reliability diagrams are increasingly reported alongside traditional metrics like accuracy, sensitivity, specificity, AUC, precision, recall, and F1 score, especially when the outcomes lead to significant actions such as biopsy, systemic therapy, or isolation protocols [74]. Evaluations are conducted in accordance with AI reporting and trial guidelines including DECIDE-AI, CLAIM, MINIMAR, STARD-AI, CONSORT-AI, SPIRIT-AI, TRIPOD-AI, PROBAST-AI, and dermatology-specific CLEAR Derm recommendations, which collectively outline how to detail cohorts, endpoints, imaging protocols, subgroup analyses, and assessments of risk of bias [43 - 52]. The consistent application of these frameworks across studies on eczema, psoriasis, melanoma, and mpox guarantees that the reported performance is interpretable, comparable, and appropriate for incorporation into precision dermatology care pathways.

Results and Translation

This section emphasizes important insights derived from six themes, which encompass the diagnostic efficacy of image-based AI in dermoscopy and clinical photography, the advantages of integrating clinical context and multimodal signals to enhance discrimination and calibration, label-efficient learning techniques that improve generalization with a reduced number of annotations, strategies to ensure reliability through explainability, uncertainty estimation, and calibration, methods that facilitate deployability and privacy, including compression, on-device inference, and federated learning, as well as study design and reporting practices that adhere to medical AI guidelines with a focus on external and subgroup-aware validation.

Diagnostic Performance of AI Methods

Across dermatology tasks, retrospective studies on curated image sets often report high discrimination, but prospective or routine-care evaluations frequently reveal performance drops due to domain shift (devices, lighting, comorbidity, and case mix). For example, pooled analyses of photo-based onychomycosis classifiers report AUCs around 0.90–0.997, yet a prospective routine-care evaluation reported a more modest AUC of 0.751 with ~70% sensitivity and ~73% specificity [2, 3]. Similar patterns motivate temporal and external validation before clinical use.

For pigmented lesions, dermoscopic melanoma models can achieve AUC values near 0.98 on benchmark data, but performance declines when models are applied to unconstrained clinical photographs and more heterogeneous lesion categories (e.g., SK/BKL) [15, 90]. A key step toward clinical actionability is prospective evaluation at the intended operating point: in a primary-care melanoma triage trial, an AI decision-support tool achieved an AUC of 0.96 (95% CI 0.928–0.980) with 95.2% sensitivity and 84.5% specificity [29]. In low-prevalence settings, such sensitivity can support 'rule-out' triage to reduce missed melanomas, but referral/biopsy

impact depends on threshold choice, baseline prevalence, and whether specificity is sufficient to avoid excessive false-positive referrals.

Inflammatory disease pipelines increasingly report component-level scoring that aligns with clinical indices. For atopic dermatitis, lesion segmentation and EASI component scoring have shown strong agreement with clinician ratings, including high erythema accuracy in controlled studies [30]. For psoriasis, PASI-related segmentation/area pipelines report moderate-to-good agreement that is promising for longitudinal severity tracking and clinical validation studies [27]. For infectious diseases such as mpox and superficial fungal infections, multiple models report high AUCs in external testing, but the evidence base remains heterogeneous and should prioritize prospective designs, calibrated risk outputs, and clearly defined downstream actions (isolation, confirmatory testing, or referral) [4, 33, 71, 72].

Multimodal and Clinical-Context Models

Fusing images with structured clinical metadata, including age, sex, anatomical location, and symptoms, can enhance accuracy and AUC, and in certain studies, calibration when compared to image-only baselines [65 - 67]. Multimodal datasets that facilitate this integration comprise HAM10000, SIIM-ISIC 2020, PAD-UFES-20, and the Derm7pt / Seven-Point Checklist benchmark, which associate dermoscopic or clinical images with fundamental patient characteristics and lesion descriptions [7 - 9, 12]. Documented shifts in domain from dermoscopy to clinical images, for instance, when models trained on HAM10000 are utilized on BCN20000, resulting in a decrease in AUC from 0.981 to 0.907 and accuracy from 82.0% to 58.8%, highlighting the necessity for metadata-aware modeling and external validation at fixed operating points [15]. Multimodal and vision-language systems such as CLIP, multimodal BiomedCLIP, and assistant-style VLMs (LLaVA-Med, Med-Flamingo, SkinGPT-4) demonstrate classification capabilities with Top-1 accuracy, AUC, precision, recall, specificity, and F1. Retrieval metrics such as Recall@K and mAP, and visual question answering (VQA) metrics including EM, BLEU, and ROUGE, are frequently supplemented by clinician-evaluated correctness and safety [40, 41, 77 - 79]. Reporting by subgroup based on skin tone and capture device guarantees that sensitivity, specificity, and AUC are evaluated within specific strata, and clarifies when the inclusion of clinical context results in significant improvements in diagnostic discrimination and calibration in practice.

Self-Supervision and Weak Labels

Self-supervised pretraining techniques such as MAE, DINOv2, and SimCLR, along with weak or text-aligned supervision through CLIP and multimodal BiomedCLIP, enhance label efficiency and external performance when annotations are scarce, demonstrating improvements in accuracy, AUC, and in certain scenarios, calibration [38 - 42, 74]. In downstream dermatology applications, Top-1 accuracy and AUC are typically reported for classification, with additional metrics such as precision, recall, specificity, and F1 being included once decision thresholds are established. Recall@K and mAP serve to measure retrieval quality in text-image contexts, thereby supporting multimodal and weak-label frameworks based on CLIP-like models [40, 41]. Thoughtful dataset design, which includes leakage-safe splits, patient- and time-level de-duplication, and balanced representation across various skin tones and capture devices, is crucial to ensure that the observed improvements in sensitivity, specificity, and AUC are maintained during external validation and across different subgroups [6 - 13, 43, 65 - 67].

Explainability and Uncertainty

Explainability and uncertainty tools enhance fundamental diagnostic metrics such as accuracy, precision, recall, specificity, F1 score, and AUC. They are frequently combined with calibration summaries (including ECE, Brier score, and reliability diagrams) to provide context for model confidence and errors [74]. Attribution techniques like Grad-CAM, Integrated Gradients, and SHAP, along with dermatology-specific saliency case studies, deliver pixel- or region-level insights that facilitate failure analysis and clinician evaluation [83 - 85, 95, 96]. Deep ensembles enhance the quality of calibration and error detection while preserving task accuracy and AUC, whereas Monte-Carlo dropout serves as a more lightweight option for selective prediction [86].

Conformal prediction provides distribution-free coverage guarantees by returning prediction sets rather than a single label [5]. In practice, studies often target 90–95% coverage and then report accuracy and AUC alongside sensitivity, specificity, and F1 score at defined accept-versus-abstain operating points. For clinical deployment, it is helpful to also report the abstention rate and the downstream action taken for abstained cases (e.g., dermatologist review or biopsy when warranted). More broadly, calibration assessments should accompany discrimination metrics: ECE and Brier score, plus sensitivity and specificity at clinically meaningful thresholds, help ensure that probability estimates are actionable for triage and patient safety. Evidence from human-AI collaboration studies suggests that clinicians supported by algorithms can achieve higher diagnostic accuracy and more favorable reader operating points than unaided clinicians, reinforcing the value of pairing explanations with calibrated uncertainty in practice [97].

Deployment and Privacy-Preserving Learning

On-device inference utilizing knowledge distillation, frequently in conjunction with methods such as INT8 quantization, facilitates model compression and acceleration without considerable loss of accuracy, while post-compression calibration and profiling guarantee appropriate latency, throughput, and energy consumption [87]. Federated learning enables multi-institutional training while maintaining data locality and is evolving from initial feasibility studies to more extensive applications in dermatology imaging [88, 89]. In a prospective primary-care melanoma triage study,

an AI decision-support tool achieved an AUC of 0.960 (95% CI 0.928-0.980), with a sensitivity of 95.2% and specificity of 84.5%, indicating readiness for clinical deployment [29]. Reader-AI analysis revealed that a CNN attained 87.5% sensitivity at 60% specificity, and at a fixed dermatologist sensitivity of 74.1%, it achieved 86.5% specificity [18, 97]. The mobile deployment of a smartphone mpox application reported an accuracy of 91.11%, showcasing effective on-device performance in practical, real-world environments [72].

Design and Reporting Evaluation

Early-stage clinical assessments and diagnostic investigations in dermatology AI must adhere to established protocols such as DECIDE-AI, CLAIM, MINIMAR, STARD-AI, CONSORT-AI, SPIRIT-AI, TRIPOD+AI, PROBAST+AI, and CLEAR Derm [43 - 52]. Furthermore, it is essential to implement transparent documentation and governance frameworks for health AI [98]. The application of segmentation with augmentation in eczema resulted in an improvement of approximately 25% in F1 and precision, with the EASI component scoring achieving an erythema accuracy of 99.17% [30]. In analyses involving melanoma reader-AI, a CNN demonstrated a sensitivity of 87.5% at a specificity of 60%, while maintaining a fixed dermatologist sensitivity of 74.1%, which corresponded to a model specificity of 86.5% [18, 97]. When transitioning from dermoscopy to clinical photographs, there was a notable decline in accuracy from 82.0% to 58.8%, and the AUC decreased from 0.981 to 0.907, underscoring the necessity for guideline-driven, external, and subgroup-stratified reporting [15, 99 - 101]. Collectively, these practices promote reproducibility, comparability, and the safe implementation of AI systems in practical dermatology.

Table 2 summarizes the main AI tasks, model families, data sources, and reported diagnostic performance metrics across inflammatory, neoplastic, and infectious skin diseases.

Figure 11 presents an illustrative example of the advantages that artificial intelligence (AI) can offer to dermatology in the context of inflammatory, neoplastic, and infectious skin conditions. In the tasks illustrated, the performance of AI typically aligns with or surpasses that of healthcare professionals, suggesting that these systems can improve diagnostic precision and reliability while aiding clinicians in their decision-making processes, rather than substituting their expertise. Although the figure does not encompass all possibilities, it acts as a representative overview of the increasing body of evidence supporting AI-assisted treatment in dermatological conditions.

In conclusion, currently, the field of Dermatology AI has reached its highest level of development in the classification of pigmented lesions. Here, models based on Convolutional Neural Networks (CNN) and Transformers demonstrate exceptional ability to differentiate between melanoma, basal cell carcinoma, melanocytic naevi, and benign keratosis-like lesions. In the context of inflammatory diseases, segmentation-to-burden pipelines yield estimates for eczema and psoriasis that align with clinician assessments, specifically through the EASI and PASI metrics. This facilitates the quantitative monitoring of disease progression and treatment efficacy using clinical photographs. Regarding infectious dermatoses, AI technologies have already exhibited significant effectiveness in diagnosing prevalent superficial fungal infections and emerging viral conditions, such as mpox. This includes the development of web- and smartphone-based applications that are appropriate for initial triage.

Clinical Outlook

Immediate and near-term clinical applications are increasingly well defined. AI for psoriasis severity tracking (and related eczema burden estimation) is ready for prospective clinical validation studies and integration into longitudinal monitoring workflows. AI-assisted pigmented-lesion triage in primary care and teledermatology has promising prospective trial data and can be deployed as a front-door prioritization tool when paired with clear referral thresholds. For infection, image-based identification of rare or emerging presentations can support teledermatology in resource-limited settings, helping route patients to appropriate testing and treatment while reducing delays. A near-future frontier is treatment guidance: models that predict or monitor treatment response (e.g., biologic response in psoriasis or immunotherapy response/toxicity risk in melanoma) from longitudinal clinical images and, where available, histopathology and multimodal biomarkers. These applications are promising but remain in an early stage and require carefully designed prospective studies.

Open Technical Challenges

Key open challenges for high-stakes precision dermatology include:

• Developing robust test-time adaptation methods to handle domain shift in real time (device, site, lighting, and population shifts).

• Creating standardized benchmarking frameworks that include temporal validation, external multi-site testing, and routine subgroup analysis (e.g., by skin tone, age, and care setting).

• Designing inherently interpretable models and clinically meaningful explanations for decisions that trigger invasive actions (e.g., biopsy) or high-risk treatments.

• Establishing calibrated uncertainty and selective-referral strategies that reliably abstain on out-of-distribution cases.

• Accounting for imperfect reference standards by developing learning and evaluation methods robust to inter-observer variability and noisy labels.

• Establishing datasets and protocols that link model explanations to clinico-path correlates and patient-relevant outcomes, improving biological and clinical validity.

• Building deployment pipelines with continuous monitoring, drift detection, and safe update mechanisms aligned with regulation and clinical governance.

In these various areas, precision dermatology is implemented through three foundational technical

components: (i) multimodal and clinical-contextual models that integrate images with structured metadata and narrative descriptions, (ii) label-efficient pretraining techniques, which encompass self-supervised and vision-language approaches, thereby ensuring strong performance even when dermatology labels are limited, and (iii) architectures tailored to specific tasks, including image-level classifiers, region-level detectors, and pixel-level segmenters that facilitate differential diagnosis, lesion identification, and severity assessment. The reliability and safety of these systems are enhanced through methods that promote explainability, predictions that account for uncertainty and are calibrated, conformal prediction techniques that guarantee coverage for output sets, and comprehensive subgroup analyses that consider variations in skin tone, device usage, and clinical environments.

Effective translation relies on infrastructure decisions: knowledge distillation and quantization for on-device inference, federated learning for multi-site collaboration without centralizing images, and deployment strategies that facilitate primary care, teledermatology, and specialist workflows. Phase-appropriate design and reporting frameworks (DECIDE-AI, CLAIM, MINIMAR, STARD-AI, CONSORT-AI, SPIRIT-AI, TRIPOD-AI, PROBAST-AI, CLEAR Derm) offer a structure for detailing cohorts, endpoints, imaging protocols, calibration, and subgroup performance, ensuring that results are interpretable and clinically actionable.

Performance is still affected by domain shifts between dermoscopy and clinical photography, across different centers, and among under-represented skin tones and devices. Consequently, priorities include curating leakage-safe, tone- and device-balanced datasets, ensuring external and temporal validation with clear subgroup reporting, continuing to invest in label-efficient and multimodal learning, compressing and enhancing privacy for models intended for edge deployment, and monitoring systems after deployment. Collectively, these measures can link benchmark improvements in AI to equitable, reliable precision-dermatology tools for inflammatory, neoplastic, and infectious skin diseases at the bedside and in routine practice.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Clinical trial registration

Not applicable.

Conflicts of interest/Competing interests

Authors declare that they have no conflicts of interest.

Availability of data and material

Not applicable.

Code availability

Not applicable.

Authors' contributions

Ihab Zaqout contributed to the conception, literature search, drafting, and final revision of the manuscript. Saeb Zakout contributed clinical expertise, critically revised the manuscript, and approved the final version. All authors approved the final version for submission.

Ethics approval

Reproduced from DermNet (https://dermnetnz.org/) with permission.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Originality Declaration for Figures

Figures 10–11 are original and have been created by the authors specifically for the purposes of this study. Figures 1–9 include representative clinical images reproduced from DermNet (https://dermnetnz.org/) with permission for educational and research purposes. No other previously published or copyrighted images have been used.

Declaration on generative AI and AI-assisted technologies in the writing process

The authors used an AI language model to assist with English-language editing and formatting only. The AI tool was not used to generate scientific content, data, results, or conclusions. The authors reviewed and edited all text and take full responsibility for the integrity and accuracy of the manuscript.

Acknowledgments

The author sincerely thanks the Council for At-Risk Academics (CARA), London, United Kingdom, and the University of Southampton, United Kingdom, for hosting the author as a visiting professor.

References

1. Alipour N, Burke T, Courtney J. Skin Type Diversity in Skin Lesion Datasets: A Review. Current Dermatology Reports. 2024 08 14;13(3):198-210. https://doi.org/10.1007/s13671-024-00440-0
2. Kim YJ, Han SS, Yang HJ, Chang SE. Prospective, comparative evaluation of a deep neural network and dermoscopy in the diagnosis of onychomycosis. PLOS ONE. 2020 06 11;15(6):e0234334. https://doi.org/10.1371/journal.pone.0234334
3. Han SS, Park GH, Lim W, Kim MS, Na JI, Park I, Chang SE. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. PLOS

ONE. 2018 01 19;13(1):e0191493. https://doi.org/10.1371/journal.pone.0191493
4. He W, Zheng C, Yang L, Tan J. Automatic fluorescence microscopic image analyzer: a novel AI-based tool for early diagnosing superficial fungal infections. BMC Infectious Diseases. 2025 07 10;25(1):908. https://doi.org/10.1186/s12879-025-11296-5
5. Angelopoulos AN, Bates S. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. 2022 Dec 07;. https://doi.org/10.48550/arXiv.2107.07511
6. Kurtansky NR, D'Alessandro BM, Gillis MC, Betz-Stablein B, Cerminara SE, Garcia R, Girundi MA, et al. The SLICE-3D dataset: 400,000 skin lesion image crops extracted from 3D TBP for skin cancer detection. Scientific Data. 2024 08 14;11(1):884. https://doi.org/10.1038/s41597-024-03743-w
7. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data. 2018 08 14;5(1):180161. https://doi.org/10.1038/sdata.2018.161
8. Rotemberg V, Kurtansky N, Betz-Stablein B, Caffery L, Chousakos E, Codella N, Combalia M, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data. 2021 01 28;8(1):34. https://doi.org/10.1038/s41597-021-00815-z
9. Pacheco AG, Lima GR, Salomão AS, Krohling B, Biral IP, De Angelo GG, Alves Jr FC, et al. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in Brief. 2020 Oct;32:106221. https://doi.org/10.1016/j.dib.2020.106221
10. Ricci Lara MA, Rodríguez Kowalczuk MV, Lisa Eliceche M, Ferraresso MG, Luna DR, Benitez SE, Mazzuoccolo LD. A dataset of skin lesion images collected in Argentina for the evaluation of AI tools in this population. Scientific data. 2023 Oct 18;10(1). https://doi.org/10.1038/s41597-023-02630-0
11. Hernández-Pérez C, Combalia M, Podlipnik S, Codella NCF, Rotemberg V, Halpern AC, Reiter O, et al. BCN20000: Dermoscopic Lesions in the Wild. Scientific Data. 2024 06 17;11(1):641. https://doi.org/10.1038/s41597-024-03387-w
12. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. 7-Point Checklist and Skin Lesion Classification using Multi-Task Multi-Modal Neural Nets. IEEE journal of biomedical and health informatics. 2018 04 09;. https://doi.org/10.1109/JBHI.2018.2824327
13. Mendonca T, Ferreira PM, Marques JS, Marcal ARS, Rozeira J. PH² - a dermoscopic image database for research and benchmarking. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2013;2013:5437-5440. https://doi.org/10.1109/EMBC.2013.6610779
14. Farooq MA, Yao W, Schukat M, Little MA, Corcoran P. Derm-T2IM: Harnessing Synthetic Skin Lesion Data via Stable Diffusion Models for Enhanced Skin Disease Classification using ViT and CNN. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2024 07;2024:1-5. https://doi.org/10.1109/EMBC53108.2024.10781852
15. Combalia M, Codella N, Rotemberg V, Carrera C, Dusza S, Gutman D, Helba B, et al. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge. The Lancet. Digital Health. 2022 05;4(5):e330-e339. https://doi.org/10.1016/S2589-7500(22)00021-8
16. Venkatesh KP, Raza MM, Nickel G, Wang S, Kvedar JC. Deep learning models across the range of skin disease. npj Digital Medicine. 2024 02 10;7(1):32, s41746-024-01033-8. https://doi.org/10.1038/s41746-024-01033-8
17. Heinlein L, Maron RC, Hekler A, Haggenmüller S, Wies C, Utikal JS, Meier F, et al. Prospective multicenter study using artificial intelligence to improve dermoscopic melanoma diagnosis in patient care. Communications Medicine. 2024 09 11;4(1):177. https://doi.org/10.1038/s43856-024-00598-5
18. Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, Razmara M, et al. Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. The Lancet. Digital Health. 2023 Oct;5(10):e679-e691. https://doi.org/10.1016/S2589-7500(23)00130-9
19. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015 03 02;. https://doi.org/10.48550/arXiv.1502.03167
20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 05 18;. https://doi.org/10.48550/arXiv.1505.04597
21. Cai L, Hou K, Zhou S. Intelligent skin lesion segmentation using deformable attention Transformer U-Net with bidirectional attention mechanism in skin cancer images. Skin research and technology: official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging. 2024 08;30(8):e13783. https://doi.org/10.1111/srt.13783
22. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, et al. Segment Anything. 2023 04 05;. https://doi.org/10.48550/arXiv.2304.02643
23. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nature Communications. 2024 01 22;15(1):654. https://doi.org/10.1038/s41467-024-44824-z
24. Ravi N, Gabeur V, Hu Y, Hu R, Ryali C, Ma T, Khedr H, et al. SAM 2: Segment Anything in Images and Videos. 2024;. https://doi.org/10.48550/ARXIV.2408.00714
25. Dong H, Gu H, Chen Y, Yang J, Chen Y, Mazurowski MA. Segment anything model 2: an application to 2D and 3D medical images. 2024;. https://doi.org/10.48550/ARXIV.2408.00756
26. Zhu J, Hamdi A, Qi Y, Jin Y, Wu J. Medical SAM 2: Segment medical images as video via Segment Anything Model 2. 2024;. https://doi.org/10.48550/ARXIV.2408.00874
27. Schaap M, Cardozo N, Patel A, De Jong E, Van Ginneken B, Seyger M. Image-based automated Psoriasis Area Severity Index scoring by Convolutional Neural Networks. Journal of the European Academy of Dermatology and Venereology. 2022 01;36(1):68-75. https://doi.org/10.1111/jdv.17711
28. Al-Waisy AS, Al-Fahdawi S, Khalaf MI, Mohammed MA, Al-Attar B, Al-Andoli MN. A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. Scientific Reports. 2025 08 25;15(1):31234. https://doi.org/10.1038/s41598-025-15655-9
29. Papachristou P, Söderholm M, Pallon J, Taloyan M, Polesie S, Paoli J, Anderson CD, Falk M. Evaluation of an artificial intelligence-based decision support for the detection of cutaneous melanoma in primary care: a prospective real-life clinical trial. The British Journal of Dermatology. 2024 06 20;191(1):125-133. https://doi.org/10.1093/bjd/ljae021
30. Medela A, Mac Carthy T, Aguilar Robles SA, Chiesa-Estomba CM, Grimalt R. Automatic SCOring of Atopic Dermatitis Using Deep Learning: A Pilot Study. JID innovations: skin science from molecules to population

health. 2022 05;2(3):100107. https://doi.org/10.1016/j.xjidi.2022.100107
31. Jeskey J, Kurien C, Blunk H, Sehmi K, Areti S, Nguyen D, Hostoffer R. Atopic Dermatitis: A Review of Diagnosis and Treatment. The journal of pediatric pharmacology and therapeutics: JPPT: the official journal of PPAG. 2024 Dec;29(6):587-603. https://doi.org/10.5863/1551-6776-29.6.587
32. Schmitt J, Langan S, Deckert S, Svensson A, Von Kobyletzki L, Thomas K, Spuls P. Assessment of clinical signs of atopic dermatitis: A systematic review and recommendation. Journal of Allergy and Clinical Immunology. 2013 Dec;132(6):1337-1347. https://doi.org/10.1016/j.jaci.2013.07.008
33. Nayak T, Chadaga K, Sampathila N, Mayrose H, Gokulkrishnan N, Bairy G M, Prabhu S, et al. Deep learning based detection of monkeypox virus using skin lesion images. Medicine in Novel Technology and Devices. 2023 06;18:100243. https://doi.org/10.1016/j.medntd.2023.100243
34. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2021 06 03;. https://doi.org/10.48550/arXiv.2010.11929
35. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. IEEE International Conference on Computer Vision. 2021;10012-22.
36. Naseer MM, Ranasinghe K, Khan SH, Hayat M, Shahbaz Khan F, Yang M. Intriguing Properties of Vision Transformers. Advances in Neural Information Processing Systems. 2021;34:23296–23308.
37. Liu Z, Mao H, Wu C, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. 2022 03 02;. https://doi.org/10.48550/arXiv.2201.03545
38. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked Autoencoders Are Scalable Vision Learners. 2021 Dec 19;. https://doi.org/10.48550/arXiv.2111.06377
39. Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, Fernandez P, et al. DINOv2: Learning Robust Visual Features without Supervision. 2024 02 02;. https://doi.org/10.48550/arXiv.2304.07193
40. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, et al. Learning Transferable Visual Models From Natural Language Supervision. 2021 02 26;. https://doi.org/10.48550/arXiv.2103.00020
41. Zhang S, Xu Y, Usuyama N, Xu H, Bagga J, Tinn R, Preston S, et al. A Multimodal Biomedical Foundation Model Trained from Fifteen Million Image–Text Pairs. NEJM AI. 2024 Dec 20;2. https://doi.org/10.1056/AIoa2400640
42. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple Framework for Contrastive Learning of Visual Representations. 2020 07 01;. https://doi.org/10.48550/arXiv.2002.05709
43. Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, Guitera P, et al. Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology: CLEAR Derm Consensus Guidelines From the International Skin Imaging Collaboration Artificial Intelligence Working Group. JAMA dermatology. 2022 01 01;158(1):90-96. https://doi.org/10.1001/jamadermatol.2021.4915
44. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, Denniston AK, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022 05 18;377:e070904. https://doi.org/10.1136/bmj-2022-070904
45. Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiology. Artificial Intelligence. 2020 03;2(2):e200029. https://doi.org/10.1148/ryai.2020200029
46. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. Journal of the American Medical Informatics Association: JAMIA. 2020 Dec 09;27(12):2011-2015. https://doi.org/10.1093/jamia/ocaa088
47. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, Golub RM, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nature Medicine. 2025 Oct;31(10):3283-3289. https://doi.org/10.1038/s41591-025-03953-8
48. Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Communications Medicine. 2024 04 11;4(1):71. https://doi.org/10.1038/s43856-024-00492-0
49. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ. 2020 09 09;370:m3164. https://doi.org/10.1136/bmj.m3164
50. Cruz Rivera S, Liu X, Chan A, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. The Lancet. Digital Health. 2020 Oct;2(10):e549-e560. https://doi.org/10.1016/S2589-7500(20)30219-3
51. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024 04 16;385:e078378. https://doi.org/10.1136/bmj-2023-078378
52. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, Beam AL, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025 03 24;388:e082505. https://doi.org/10.1136/bmj-2024-082505
53. Yilmaz A, Yasar SP, Gencoglan G, Temelkuran B. DERM12345: A Large, Multisource Dermatoscopic Skin Lesion Dataset with 40 Subclasses. Scientific Data. 2024 Nov 28;11(1):1302. https://doi.org/10.1038/s41597-024-04104-3
54. Li H, Chen G, Zhang L, Xu C, Wen J. A review of psoriasis image analysis based on machine learning. Frontiers in Medicine. 2024 08 07;11:1414582. https://doi.org/10.3389/fmed.2024.1414582
55. Luz F. Silva D, Rubinho R, Denofre A, Avila S, Magalhães RF. Artificial intelligence on inflammatory dermatoses: where are we and where are we going?. Anais Brasileiros de Dermatologia. 2025 09;100(5):501164. https://doi.org/10.1016/j.abd.2025.501164
56. Tang A, Wei ML, Haemel A, La C, Sirota M, Lee EY. Artificial intelligence-enabled precision medicine for inflammatory skin diseases. ArXiv. 2025 05 14;:arXiv:2505.09527v1.
57. Uwishema O, Ghezzawi M, Charbel N, Alawieh S, Roy S, Wojtara M, Hakayuwa CM, et al. Diagnostic performance of artificial intelligence for dermatological conditions: a systematic review focused on low- and middle-income countries to address resource constraints and improve access to specialist care. International Journal of Emergency Medicine. 2025 09 29;18(1):172. https://doi.org/10.1186/s12245-025-00975-4
58. Brancaccio G, Balato A, Malvehy J, Puig S, Argenziano G,

Kittler H. Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check. Journal of Investigative Dermatology. 2024 03;144(3):492-499. https://doi.org/10.1016/j.jid.2023.10.004
59. Biswas S, Achar U, Hakim B, Achar A. Artificial Intelligence in Dermatology: A Systematized Review. International Journal of Dermatology and Venereology. 2025 03;8(1):33-39. https://doi.org/10.1097/JD9.0000000000000404
60. Zbrzezny AM, Krzywicki T. Artificial Intelligence in Dermatology: A Review of Methods, Clinical Applications, and Perspectives. Applied Sciences. 2025 07 14;15(14):7856. https://doi.org/10.3390/app15147856
61. DermNet New Zealand Trust. Dermatology across Inflammation, Cancer, and Infection [Internet]. Auckland: DermNet NZ; 1995. [cited 2025 Aug 11]. Available from: https://dermnetnz.org/.
62. Huang K, Wu X, Li Y, Lv C, Yan Y, Wu Z, Zhang M, et al. Artificial Intelligence-Based Psoriasis Severity Assessment: Real-world Study and Application. Journal of Medical Internet Research. 2023 03 16;25:e44932. https://doi.org/10.2196/44932
63. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology: Official Journal of the European Society for Medical Oncology. 2018 08 01;29(8):1836-1842. https://doi.org/10.1093/annonc/mdy166
64. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015 Dec 10;. https://doi.org/10.48550/arXiv.1512.03385
65. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, Ko J, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Science Advances. 2022 08 12;8(32):eabq6147. https://doi.org/10.1126/sciadv.abq6147
66. Groh M, Harris C, Soenksen L, Lau F, Han R, Kim A, Koochek A, Badri O. Evaluating Deep Neural Networks Trained on Clinical Images in Dermatology with the Fitzpatrick 17k Dataset. arXiv.org. 2021 04 20;.
67. Pacheco AGC, Krohling RA. The impact of patient clinical information on automated skin cancer detection. Computers in Biology and Medicine. 2020 01;116:103545. https://doi.org/10.1016/j.compbiomed.2019.103545
68. Comes MC, Fucci L, Mele F, Bove S, Cristofaro C, De Risi I, Fanizzi A, et al. A deep learning model based on whole slide images to predict disease-free survival in cutaneous melanoma patients. Scientific Reports. 2022 Nov 27;12(1):20366. https://doi.org/10.1038/s41598-022-24315-1
69. Manzoor K, Gilal NU, Agus M, Schneider J. Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. DIGITAL HEALTH. 2025 05;11:20552076251351858. https://doi.org/10.1177/20552076251351858
70. García-Rayo MC, Avilés-Izquierdo JA, García-Piqueras P, Parra-Blanco V, Rodríguez Lomba E. Clues on dermoscopy. Seborrheic keratoses: when the good ones look bad:. European Journal of Dermatology. 2022 Nov 01;32(6):814-815. https://doi.org/10.1684/ejd.2022.4385
71. Thieme AH, Zheng Y, Machiraju G, Sadee C, Mittermaier M, Gertler M, Salinas JL, et al. A deep-learning algorithm to classify skin lesions from mpox virus infection. Nature Medicine. 2023 03;29(3):738-747. https://doi.org/10.1038/s41591-023-02225-7
72. Sahin VH, Oztel I, Yolcu Oztel G. Human Monkeypox Classification from Skin Lesion Images with Deep Pre-trained Network using Mobile Application. Journal of Medical Systems. 2022 Oct 10;46(11):79. https://doi.org/10.1007/s10916-022-01863-7
73. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv.org. 2019 05 28.
74. Minderer M, Djolonga J, Romijnders R, Hubis F, Zhai X, Houlsby N, Tran D, Lucic M. Revisiting the Calibration of Modern Neural Networks. arXiv.org. 2021 06 15.
75. Lin T, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. arXiv.org. 2017 08 0.
76. Bochkovskiy A, Wang C, Liao HM. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv.org. 2020 04 23.
77. Li C, Wong C, Zhang S, Usuyama N, Liu H, Yang J. LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day. Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023;1240:28541 – 64. https://doi.org/10.5555/3666122.3667362.
78. Zhou J, He X, Sun L, Xu J, Chen X, Chu Y. Pre-trained multimodal large language model enhances dermatological diagnosis.
79. Moor M, Huang Q, Wu S, Yasunaga M, Zakka C, Dalmia Y, Reis EP, Rajpurkar P, Leskovec J. Med-Flamingo: a Multimodal Medical Few-shot Learner. arXiv.org. 2023 07 27.
80. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. 2019 01 04;. https://doi.org/10.48550/arXiv.1711.05101
81. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2017 01 30;. https://doi.org/10.48550/arXiv.1412.6980
82. Pal M, Mahal A, Mohapatra RK, Obaidullah AJ, Sahoo RN, Pattnaik G, Pattanaik S, et al. Deep and Transfer Learning Approaches for Automated Early Detection of Monkeypox (Mpox) Alongside Other Similar Skin Lesions and Their Classification. ACS omega. 2023 09 05;8(35):31747-31757. https://doi.org/10.1021/acsomega.3c02784
83. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. International Journal of Computer Vision. 2020 02;128(2):336-359. https://doi.org/10.1007/s11263-019-01228-7
84. Sundararajan M, Taly A, Yan Q. Axiomatic Attribution for Deep Networks. 2017 06 13;. https://doi.org/10.48550/arXiv.1703.01365
85. Lundberg S, Lee S. A Unified Approach to Interpreting Model Predictions. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;4768:– 77. https://doi.org/10.5555/3295222.3295230.
86. Faghani S, Moassefi M, Rouzrokh P, Khosravi B, Baffour F, Ringler M. Quantifying Uncertainty in Deep Learning of Radiologic Images. Radiology. 2023;308(2). https://doi.org/10.1148/radiol.222217.
87. Moslemi A, Briskina A, Dang Z, Li J. A survey on knowledge distillation: Recent advancements. Machine Learning with Applications. 2024 Dec 01;18:100605. https://doi.org/10.1016/j.mlwa.2024.100605
88. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, Milchenko M, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports. 2020 07 28;10(1):12598. https://doi.org/10.1038/s41598-020-69250-1
89. Rehman MHU, Hugo Lopez Pinaya W, Nachev P, Teo JT, Ourselin S, Cardoso MJ. Federated learning for medical imaging radiology. The British Journal of Radiology.

2023 Oct;96(1150):20220890. https://doi.org/10.1259/bjr.20220890
90. Sarı MO, Keser K. Classification of skin diseases with deep learning based approaches. Scientific Reports. 2025 07 28;15(1):27506. https://doi.org/10.1038/s41598-025-13275-x
91. Hossain MS, Ahmed M, Rahman MS. From survey to solution: A deep learning framework for reliable monkeypox diagnosis using skin images. Array. 2025 Dec;28:100554. https://doi.org/10.1016/j.array.2025.100554
92. Cao Y, Yue Y, Ma X, Liu D, Ni R, Liang H, Li Z. Robustly detecting mpox and non-mpox using a deep learning framework based on image inpainting. Scientific Reports. 2025 01 10;15(1):1576. https://doi.org/10.1038/s41598-025-85771-z
93. Uysal F. Detection of Monkeypox Disease from Human Skin Images with a Hybrid Deep Learning Model. Diagnostics. 2023 05 17;13(10):1772. https://doi.org/10.3390/diagnostics13101772
94. Sorayaie Azar A, Naemi A, Babaei Rikan S, Bagherzadeh Mohasefi J, Pirnejad H, Wiil UK. Monkeypox detection using deep neural networks. BMC Infectious Diseases. 2023 06 27;23(1):438. https://doi.org/10.1186/s12879-023-08408-4
95. Muhammad D, Bendechache M. Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis. Computational and Structural Biotechnology Journal. 2024 Dec;24:542-560. https://doi.org/10.1016/j.csbj.2024.08.005
96. Chen H, Gomez C, Huang C, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. npj Digital Medicine. 2022 Oct 19;5(1):156. https://doi.org/10.1038/s41746-022-00699-2
97. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, et al. Human–computer collaboration for skin cancer recognition. Nature Medicine. 2020 08;26(8):1229-1234. https://doi.org/10.1038/s41591-020-0942-0
98. Gilbert S, Adler R, Holoyad T, Weicken E. Could transparent model cards with layered accessible information drive trust and safety in health AI?. npj Digital Medicine. 2025 02 25;8(1):124. https://doi.org/10.1038/s41746-025-01482-9
99. Naseri H, Safaei AA. Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: a systematic literature review. BMC Cancer. 2025 01 13;25(1):75. https://doi.org/10.1186/s12885-024-13423-y
100. Winkler JK, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Thomas L, et al. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. JAMA dermatology. 2019 Oct 01;155(10):1135-1141. https://doi.org/10.1001/jamadermatol.2019.1735
101. Winkler JK, Sies K, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, et al. Melanoma recognition by a deep learning convolutional neural network-Performance in different melanoma subtypes and localisations. European Journal of Cancer. 2020 03;127:21-29. https://doi.org/10.1016/j.ejca.2019.11.020